

Attaining proficiency with endobronchial ultrasound-guided transbronchial needle aspiration

Yinin Hu, MD,^a Varun Puri, MD,^b Traves D. Crabtree, MD,^b Daniel Kreisel, MD, PhD,^b Alexander S. Krupnick, MD,^b Alexander G. Patterson, MD,^b and Bryan F. Meyers, MD, MPH^b

Objectives: Endobronchial ultrasound-guided transbronchial needle aspiration (EBUS-TBNA) is becoming the preferred method of mediastinal staging for lung cancer. We investigated the learning curve for EBUS-TBNA using risk-adjusted cumulative sum (Cusum).

Methods: A retrospective study of EBUS-TBNA was performed at a single academic institution for patients with mediastinal or hilar lymphadenopathy in the setting of proven or suspected lung cancer. A sampling pass was defined as a full retraction and repositioning of the aspiration needle. Rapid on-site evaluation was not available. To track proficiency, risk-adjusted Cusum analysis was performed using acceptable and unacceptable failure rates of 10% and 20%, respectively. Failure was defined as false negative or nondiagnostic results.

Results: During the study period, 231 patients underwent EBUS-TBNA. Prevalence of mediastinal or hilar malignancy was 66.7% (154 out of 231). Sensitivity was 92.2% (142 out of 154), and negative predictive value was 87.9% (58 out of 66). Node size was identified as a significant predictor of EBUS-TBNA success by multiple regression. Risk-adjusted Cusum analysis demonstrated that the first and only unacceptable decision interval was crossed at 22 cases. Individual practitioner learning curves were highly variable, and the operator with the highest volume was the most consistently proficient.

Conclusions: In our experience, attainment of an acceptable failure rate for EBUS-TBNA required 22 cases. Node size is a predictor of EBUS-TBNA success. Risk-adjusted Cusum proved a powerful evaluative tool to monitor the training process of this new procedure. (*J Thorac Cardiovasc Surg* 2013;146:1387-92)

 Supplemental material is available online.



Earn CME credits at
<http://cme.ctsnetjournals.org>

Accurate mediastinal staging of patients with lung cancer is critical for therapeutic decision making and prognosis.¹ In most surgical series, pathologic staging with mediastinoscopy has been the gold standard in preoperative evaluation of mediastinal lymphadenopathy, with large clinical studies demonstrating good sensitivity and low morbidity.² However, mediastinoscopy has drawbacks, including the need for

general anesthesia, its invasive nature, potential for complications, and the inability to evaluate hilar and inferior mediastinal node stations. When applied to patients with suspected lung cancer and radiographic evidence of mediastinal lymphadenopathy, the accuracy of endobronchial ultrasound-guided transbronchial needle aspiration (EBUS-TBNA) is comparable to mediastinoscopy with an expected sensitivity of 90% or greater.^{3,4} When used in conjunction with endoscopic ultrasound, it also allows the pathologic staging of almost all mediastinal node stations.

Enthusiasm for the EBUS-TBNA procedure has driven many physicians to incorporate this staging modality into their practices. Unfortunately, the Halstedian apprenticeship model is not feasible for the majority of established practitioners who desire training in EBUS-TBNA, and there are no current requirements mandating bronchoscopic training before application in patients.⁵ To develop and maintain proficiency with EBUS-TBNA, an evaluative tool is necessary that can measure proficiency during the training period and beyond. Cumulative sum (Cusum) is one such tool that compares real-world performance to a predetermined definition of proficiency. In medical training, Cusum has successfully been applied to procedures such as placement of epidural catheters, sentinel lymph node biopsy, and thoracoscopic thymectomy.⁶⁻⁸ Here, we describe the application of Cusum analysis to evaluate the learning curve for EBUS-TBNA. Our objective is to establish the merits of

From the Department of Surgery,^a University of Virginia School of Medicine, Charlottesville, Va; and Department of Thoracic Surgery,^b Washington University School of Medicine, St Louis, Mo.

Disclosures: Authors have nothing to disclose with regard to commercial support.

Y. H. and V. P. contributed equally to this work.

Received for publication Feb 17, 2013; revisions received July 1, 2013; accepted for publication July 26, 2013; available ahead of print Sept 26, 2013.

Address for reprints: Yinin Hu, MD, Department of Surgery, University of Virginia School of Medicine, PO Box 800679, Charlottesville, VA 22903-0679 (E-mail: yh9b@virginia.edu).

0022-5223/\$36.00

Copyright © 2013 by The American Association for Thoracic Surgery

<http://dx.doi.org/10.1016/j.jtcvs.2013.07.077>

Abbreviations and Acronyms

CUSUM	= cumulative sum
EBUS-TBNA	= endobronchial ultrasound-guided transbronchial needle aspiration
ROSE	= rapid on-site pathologic evaluation

Cusum analysis for the purpose of monitoring the adoption of EBUS-TBNA at institutional and individual levels.

METHODS

All patients with known or suspected lung cancer undergoing EBUS-TBNA for tissue diagnosis or staging between January 2007 and October 2010 at the Washington University in St Louis School of Medicine, St Louis, Mo, were prospectively entered into a database. Tissue diagnosis procedures are defined as those for patients with radiologic evidence of unresectable malignant disease who received EBUS-TBNA to obtain tissue for pathologic diagnosis. Staging procedures are those performed for patients with potentially resectable disease. Preceding chart review, data collection, and analysis, the full study protocol underwent approval by the Institutional Review Board of the Washington University in St Louis School of Medicine. Patient demographics, clinical and radiologic staging information, EBUS-TBNA details, subsequent procedure details, pathology results, and clinical outcomes were retrospectively obtained via electronic chart review. A total of 254 patients were reviewed. Patients receiving a negative or nondiagnostic EBUS-TBNA who failed to follow-up with additional tissue sampling or radiographic surveillance were excluded (23 out of 254; 9.1%), for a final study cohort of 231. Within the study group of 231 patients, mean age was 62.5 years, and 118 out of 231 participants (51%) were men (Table 1).

A positive EBUS-TBNA was defined as pathology results consistent with malignancy or benign nodal disease; that is, histoplasmosis, sarcoidosis, or necrotizing granuloma. A negative result was defined as normal lymphoid findings or reactive lymphadenopathy. A procedure was considered nondiagnostic if it failed to produce adequate sampling, or if the sample yielded indeterminate results. Negative or nondiagnostic results from EBUS-TBNA were followed by mediastinoscopy or surgical resection, or were followed by repeat computed tomography imaging at an interval of 6 months to evaluate for mediastinal node progression. All EBUS-TBNA cytology samples diagnostic of malignancy or benign disease were assumed to be true positives. False negatives were defined as cases of nondiagnostic or negative EBUS-TBNA in which the final surgical node stage was N1 or greater, or cases in which the patient had evidence of mediastinal disease progression on follow-up imaging.

All cases of EBUS-TBNA considered for this study were performed by thoracic surgeons under general anesthesia using a linear endobronchial ultrasound scope. None of the participant surgeons had prior EBUS-TBNA experience before the study period, and no participant received formal training for the procedure. Rapid on-site pathologic evaluation (ROSE) of biopsy specimens was not routinely performed. Selective EBUS-TBNA sampling was guided by preoperative radiographic staging. Mediastinal lymph nodes subject to sampling by EBUS-TBNA included those >1 cm on preoperative imaging or during endobronchial ultrasound. A single aspiration, or "pass," of a lymph node was defined as any number of sampling oscillations with the biopsy needle along a single axis.

Cusum analysis for depiction of learning progression is described in detail elsewhere by Bolsin and Colson.⁹ Briefly, a classic Cusum analysis evokes trainer-defined parameters to measure a trainee's proficiency at an assigned task, and iterates this measurement for subsequent repetitions. Measurement of proficiency is based on a binary outcome for each

performance of a given task (success vs failure). The trainer determines a priori acceptable and unacceptable failure rates (p_0 and p_1 , respectively), which derive a numeric decrement (s) representing each success and increment (1-s) representing each failure, based on the following calculation:

$$s = \ln[(1-p_0)/(1-p_1)] / \{\ln(p_1/p_0) + \ln[(1-p_0)/(1-p_1)]\}$$

Graphic depiction of Cusum of all deflections depicts the classic learning curve. By defining type 1 and type 2 error rates, the trainer derives acceptability/unacceptability boundaries that demarcate when a trainee has crossed into proficiency or inadequacy. A type 1 error (α) is the wrongful accusation of inadequacy, whereas a type 2 error (β) is the wrongful certification of proficiency. For ease of graphic interpretation, acceptable α and β are set to be equal. The acceptability/unacceptability boundary spacing (h_0) is then determined by the following calculation:

$$h_0 = \ln[(1-\alpha)/\beta] / \{\ln(p_1/p_0) + \ln[(1-p_0)/(1-p_1)]\}$$

Thus, a Cusum curve that trends upward and crosses a series of unacceptability lines depicts a trainee who is inadequate, whereas a curve that trends downward or maintains within the bounds of 2 acceptability lines depicts a trainee who is proficient (Figure 1).

Cusum calculation adjustments for risk are discussed thoroughly by Steiner and colleagues.¹⁰ Case-specific risk factors are identified through multiple regression and used to modify the increments and decrements associated with failure and success, respectively. For example, for patient t with risk of failure q_t , when the odds ratio of failure for proficiency is set to R_0 and odds ratio of failure for inadequacy is set to R_1 , the deflections become modified to the following:

$$\ln[(1-q_t+R_0q_t)/(1-q_t+R_1q_t)] \text{ success}$$

$$\ln\{[(1-q_t+R_0q_t)R_1]/[(1-q_t+R_1q_t)R_0]\} \text{ failure}$$

For the purposes of our study, a successful EBUS-TBNA was defined as a true positive or a true negative procedure result. A failed EBUS-TBNA was defined as a nondiagnostic or false negative result. Values for acceptable and unacceptable failure rates as well as type 1 and type 2 error rates were determined by expert consensus within our institution and from literature review. Because all nondiagnostic or negative EBUS-TBNA's receive pathologic verification or subsequent follow-up, the risks of a failed procedure are mild. The linear EBUS provides a view of mediastinal anatomy foreign to most new practitioners, and the procedure was considered moderate in difficulty. Given that literature consensus on sensitivity of EBUS-TBNA is roughly 90%,¹¹ an acceptable failure rate was defined as $p_0 = 0.1$, whereas an unacceptable rate was defined as $p_1 = 0.2$. Type 1 and type 2 errors were set to be equivalent at $\alpha = \beta = 0.1$. Cusum curves were generated for our institution as a whole and for individual surgeons who performed a minimum of 20 cases during the study period. Risk-adjusted Cusum was calculated on an institution level based on significant predictors of procedure success as determined by multiple logistic regression of contributing factors node size, tissue-sampling versus staging cohort, and number of nodes sampled. The primary outcomes of our study were numbers of cases necessary to attain proficiency on an institution level based on unadjusted and risk-adjusted Cusum analyses. Secondary outcomes included Cusum results of individual practitioners and significant predictors of procedural success by logistic multiple regression.

RESULTS

EBUS-TBNA was performed for tissue diagnostic purposes for 114 patients (49.4%), and for staging for all others. Distribution of disease included lung cancer, metastatic disease, lymphoma, and benign disorders such as sarcoidosis, histoplasmosis, and necrotizing granuloma.

TABLE 1. Demographics and disease prevalence

Variable	Result
Patient demographic	
Total, n	231
Age, y	62.7 ± 12.6
Women	113 (48.9)
Disease prevalence	
Non-small cell lung cancer	132 (57.1)
Small cell lung cancer	31 (13.4)
Metastasis	27 (11.7)
Lymphoma	8 (3.5)
Benign*	32 (13.9)

Values are presented as mean ± standard deviation or n (%). *Benign diseases include sarcoidosis, fungal infection, and necrotizing granuloma.

The final disease prevalence for nodal disease of any kind was 66.7% (154 out of 231). EBUS-TBNA yielded adequate sampling for pathologic diagnosis in 90% (201 out of 231) of cases, with an overall sensitivity of 92.2% (142 out of 154), accuracy of 86.6% (200 out of 231) and negative predictive value of 87.9% (58 out of 66).

To determine the relationship between EBUS-TBNA sensitivity and thoroughness of node sampling, procedures were categorized based on the number of passes performed per node and number of node stations sampled per procedure. An average of 3.36 passes were made to each node sampled. Sensitivity of EBUS-TBNA did not improve further among cases involving >3 passes per node. Sensitivity was highest at 94.1% (48 out of 51) when lymph nodes were sampled with three passes, and lowest at 90.7% (39 out of 43) among procedures for which the number of passes was not recorded. Average number of node stations sampled per procedure was 1.51, and was not significantly different between the staging and tissue-diagnosis cohorts (1.50 vs 1.51, respectively). Sensitivity of EBUS-TBNA was not significantly correlated with number of node stations sampled.

Cumulative observed sensitivity of EBUS-TBNA was compared with an expected sensitivity of 90% (Figure E1). After an initial period of high volatility due to low volume, cumulative sensitivity approximated

expected levels beyond 90 cases. Unadjusted Cusum analysis was performed implementing an acceptable failure rate of 0.1, an unacceptable failure rate of 0.2, and equivalent type 1 and type 2 error rates of 0.1. At an institution level, the first and only unacceptable threshold was crossed at 32 cases, indicating maintenance of a proficient level of practice beyond this case number. Logistical regression using predictors node size, number of nodes sampled, and staging versus tissue diagnosis cohort revealed that only node size was a significant predictor of procedure success ($P = .044$). Under risk-adjusted Cusum analysis accounting for node size, the first and only unacceptable threshold was crossed at 22 cases, indicating proficiency thereafter (Figure 2). All procedures were performed by thoracic surgery faculty with or without the assistance of thoracic fellows. Five practitioners performed an average of 45 cases, with case volume ranging from 7 to 116 at the time of analysis. Cusum analyses for individuals were performed for the three busiest operators at our institution. Comparison of Cusum curves of individual surgeons demonstrated variability in the rate of acquiring proficiency (Figure 3).

DISCUSSION

Current guidelines for surgical training are often set with the assumption that repetition leads to proficiency. Rather than using objective measures of ability, trainees are considered proficient after a target number of attempts. However, learning curves among trainees differ greatly; a recent survey of graduating pulmonary fellows revealed that 50% believed that their trans-bronchial needle aspiration training was inadequate.¹² Although the American College of Chest Physicians recommends performing 50 radial EBUS procedures to attain a level of proficiency,¹³ this number may underestimate or overestimate the actual volume necessary among many trainees,¹⁴ and does not address differences between radial and linear EBUS. Intermittent testing using standardized assessment tools such as the Bronchoscopy Skills and Tasks Assessment Tool have demonstrated good internal and external

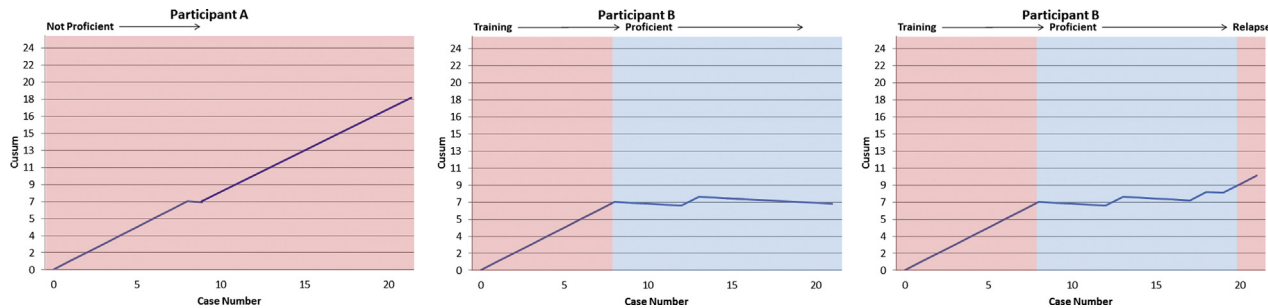


FIGURE 1. Example risk-adjusted cumulative sum (Cusum) graphs. Positive deflections indicate failed attempts, negative deflections indicate successful attempts. Horizontal lines demarcate unacceptable and acceptable thresholds. From left to right, graphs depict a subpar performer, a performer in the process of training, and a learner reverting to inadequacy after a period of proficiency.

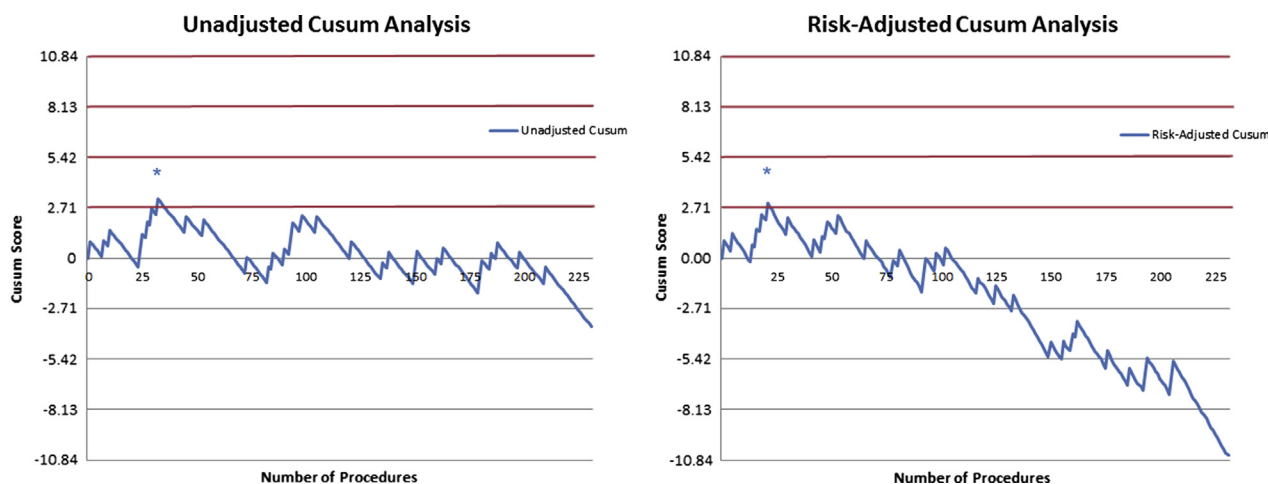


FIGURE 2. Institutional risk-adjusted cumulative sum (Cusum). *Left*, Unadjusted Cusum shows that an unacceptable threshold (red lines) was crossed at 32 cases (indicated by *), and only acceptable thresholds (gray lines) are crossed in subsequent cases, indicating proficient performance beyond 32 cases. *Right*, Risk-adjusted Cusum shows an unacceptable threshold crossed at 22 cases (indicated by *) with subsequent proficiency.

validity,¹⁵ but do not have direct application to rate of success in a clinical setting.

To our knowledge, our study is the first to apply risk-adjusted Cusum analysis to a department-wide adoption of EBUS-TBNA using strict radiographic staging criteria for sampling and without the assistance of ROSE. Our findings suggest that risk-adjusted Cusum analysis may be used to track acquisition of proficiency with EBUS-TBNA as an institution, which in our series occurred after 22 cases. This result stands in contrast to a comparison of expected and

observed cumulative sensitivity. Cumulative sensitivity is demonstrably inferior to Cusum for tracking training progress given its inability to incorporate risk-modification and reflect point-in-time performance. Predictive factors for success of EBUS-TBNA included the size of lymph nodes sampled. Individual practitioners' learning curves were highly variable within our institution, and highlighted the need for continuation of outcomes-based measures of quality. Neither number of passes per lymph node nor number of lymph nodes sampled was associated with EBUS-TBNA success.

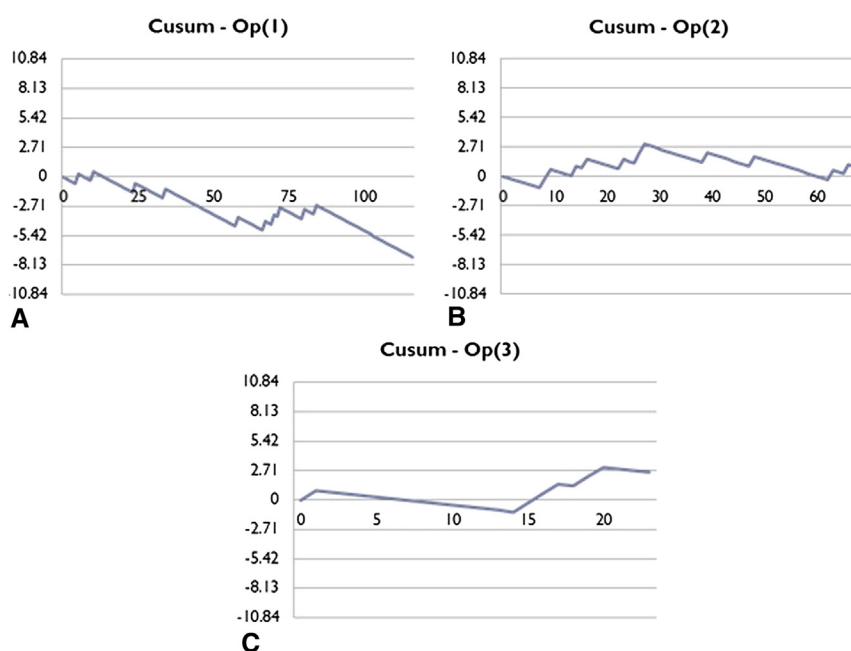


FIGURE 3. Individual risk-adjusted cumulative sum (Cusum) curves. A, Operator 1 demonstrated proficiency throughout the study period. B, Operator 2 crosses an unacceptable threshold at 27 cases, and then maintains proficiency thereafter. C, Operator 3 has yet to demonstrate proficiency after 22 cases.

To date, several studies have attempted to describe the learning curve of EBUS-TBNA in real-world settings. Tracking of proficiency surrogates such as sensitivity and diagnostic rate in a clinical setting may produce a graphically meaningful learning curve. However, these measures fall short when attempting to delineate a threshold of proficiency, because any statistical comparison of serial measures of yield requires division of patients into arbitrary temporal cohorts.¹⁶ It is encouraging, nevertheless, that a recent study by Abu-Hijleh and colleagues¹⁷ tracking serial EBUS-TBNA diagnostic yield over cohorts of 25 patients found a jump in sensitivity following the first 25 to 50 cases. This lends external validity to the results of our Cusum analysis.

During the past 20 years, Cusum has been applied in the surgical field to detect small deviations from expected outcomes for established, high-risk procedures.¹⁸ Over time, as investigators grew savvy to the inability of classic Cusum to address case variability and patient risk factors, more sophisticated, risk-adjusted Cusum methodologies were developed and extensively studied.¹⁹⁻²² Although Cusum analysis for medical procedures has been used with success in several fields, its application to bronchoscopy is sparse. On an individual level, mastery of EBUS-TBNA is a complex and highly variable process.²³ A multicenter retrospective study by Kemp and colleagues²⁴ applied unadjusted Cusum to several individual practitioners at different institutions with variable criteria for node sampling and elucidated significant differences in learning curves. This is in keeping with the unsettling variability among trainees when relating volume to proficiency. Thus far, no individual surgeon has required retraining once proficiency was obtained. Looking forward, we anticipate that continued monitoring of individual performance with Cusum analyses after every 20 cases will promote quality assurance. One weakness of classic, unadjusted Cusum is its inability to account for case variability in its assessment of performance. By incorporating risk adjustment for variables predictive of success, Cusum becomes a more powerful and valid evaluative tool.²⁵ Within our series, adjusting for node size in Cusum calculations allowed amendment of our institutional time-to-proficiency.

Studying the learning curve of an entire department has several merits. First, the installment of a new biopsy technique depends not only on the proficiency of physicians, but also the support staff in the operating room, surgical pathologists, and the hospital infrastructure. Second, modeling the learning curve of a department accounts for assistance and teaching between colleagues. Last, institution-wide application of Cusum allows identification of the need for infrastructure retraining and protocol adjustments in addition to individual retraining. For example, deviation across multiple unacceptability thresholds may

prompt a quality investigation that elucidates the need for adoption of ROSE.

Our experience with EBUS-TBNA thus far does not include the routine use of ROSE. Aspirated samples are subjectively studied by a cytotechnologist before submission for cytology and pathology. In our experience, lymph node sampling with >3 passes was not associated with increased sensitivity or negative predictive value. This is in concordance with the study by Abu-Hijleh and colleagues¹⁷ showing no increase in diagnostic yield for cases necessitating >3 aspirations per node. We did find a significant correlation between node size and EBUS-TBNA success, and this may be due to relative ease of the procedure and higher disease prevalence.

Our study is not without limitations. First, as a retrospective, single-institution study, there are inherent limitations when attempting to apply our results to influence technical protocols at other institutions. For example, the number of passes and node stations necessary to procure sufficient tissue could be affected by the availability of ROSE, which we did not employ. Second, the guidelines followed by study participants for mediastinal sampling were equivalent to “selective sampling” as defined by Detterbeck and colleagues,²⁶ and it is possible that employment of “complete sampling” or “systematic sampling” protocols could alter the expected success rate—and thus the learning curve—of EBUS-TBNA. However, differences in technical protocol would not affect the applicability of Cusum as a learning assessment tool at most centers. Additionally, our risk-adjusted Cusum analysis accommodated for node size as the sole predictive variable in relation to procedure success. With higher volume for analysis, it is possible that additional risk factors may be elucidated through multiple regression. Lastly, our results pertain to the experience of a high-volume academic center with multiple collaborative practitioners of EBUS-TBNA. Thus, our institution-level results may not be applicable to a smaller practice with fewer opportunities for collaboration. Because the establishment of a high-quality EBUS-TBNA service is clearly a multidisciplinary effort, the learning curve for an institution adopting this technique will also vary based on the ancillary resources available.

CONCLUSIONS

EBUS-TBNA is an accurate evaluative tool for mediastinal adenopathy with advantages of decreased morbidity and wider lymph node accessibility over mediastinoscopy. Risk-adjusted Cusum analysis allows real-time monitoring of proficiency levels and provides strict criteria for retraining. Institution-wide Cusum analysis adjusting for node size accurately delineates time to proficiency, which at a large academic center is approximately 22 cases. We encourage further exploration of Cusum analysis as a widely

applicable evaluative tool for trainees learning new surgical techniques.

References

1. Detterbeck FC, Boffa DJ, Tanoue LT. The new lung cancer staging system. *Chest*. 2009;136:260-71.
2. Detterbeck FC, Jantz MA, Wallace M, Vansteenkiste J, Silvestri GA. Invasive mediastinal staging of lung cancer: ACCP evidence-based clinical practice guidelines (2nd edition). *Chest*. 2007;132:202-20.
3. Herth FJ, Eberhardt R, Vilman P, Krasnik M, Ernst A. Real-time endobronchial ultrasound guided transbronchial needle aspiration for sampling mediastinal lymph nodes. *Thorax*. 2006;61:795-8.
4. Yasufuku K, Chiyo M, Koh E, Moriya Y, Iyoda A, Sekine Y, et al. Endobronchial ultrasound guided transbronchial needle aspiration for staging of lung cancer. *Lung Cancer*. 2005;50:347-54.
5. Epstein RM. Assessment in medical education. *N Engl J Med*. 2007;356:387-96.
6. Naik VN, Devito I, Halpern SH. Cusum analysis is a useful tool to assess resident proficiency at insertion of labour epidurals. *Can J Anaesth*. 2003;50:694-8.
7. Toker A, Tanju S, Ziyade S, Kaya S, Dilege S. Learning curve in videothoracoscopic thymectomy: how many operations and which situations? *Eur J Cardiothorac Surg*. 2008;34:155-8.
8. East JM, Valentine CS, Kanchev E, Blake GO. Sentinel lymph node biopsy for breast cancer using methylene blue dye manifests a short learning curve among experienced surgeons: a prospective tabular cumulative sum analysis. *BMC Surg*. 2009;9:2.
9. Bolsin S, Colson M. The use of the Cusum Technique in the assessment of trainee competence in new procedures. *Int J Qual Health Care*. 2000;12:433-8.
10. Steiner SH, Cook RJ, Farewell VT, Treasure T. Monitoring surgical performance using risk-adjusted cumulative sum charts. *Biostatistics*. 2000;1:441-52.
11. Adams K, Shah PL, Edmonds L. Test performance of endobronchial ultrasound and transbronchial needle aspiration biopsy for mediastinal staging in patients with lung cancer: a systematic review and meta-analysis. *Thorax*. 2009;64:757-62.
12. Patis NJ, Nietert PJ, Silvestri GA. Variation in training for interventional pulmonary procedures among US pulmonary/critical care fellowships: a survey of fellowship directors. *Chest*. 2005;127:1614-21.
13. Ernst A, Silvestri GA, Johnstone D. Interventional pulmonary procedures: guidelines from the American College of Chest Physicians. *Chest*. 2003;123:1693-717.
14. Steinfert DP, Hew MJ, Irving LB. Bronchoscopic evaluation of the mediastinum using endobronchial ultrasound: description of the first 216 cases carried out at an Australian tertiary hospital. *Intern Med J*. 2011;41:815-24.
15. Davoudi M, Osann K, Colt HG. Validation of two instruments to assess technical bronchoscopic skill using virtual reality simulation. *Respiration*. 2008;76:92-101.
16. Groth SS, Whitson BA, D'Cunha J, Maddaus MA, Alsharif M, Andrade RS. Endobronchial ultrasound-guided fine-needle aspiration of mediastinal lymph nodes: a single institution's early learning curve. *Ann Thorac Surg*. 2008;86:1104-9.
17. Abu-Hijleh M, El-Sameed Y, Eldridge K, Vadia E, Chiu H, Dreyfuss Z, et al. Linear probe endobronchial ultrasound bronchoscopy with guided transbronchial needle aspiration (EBUS-TBNA) in the evaluation of mediastinal and hilar pathology: introducing the procedure to a teaching institution. *Lung*. 2013;191:109-15.
18. de Leval MR, Francois K, Bull C, Brawn W, Spiegelhalter D. Analysis of a cluster of surgical failures. Application to a series of neonatal arterial switch operations. *J Thoracic Cardiovasc Surg*. 1994;107:914-24.
19. Spiegelhalter D, Grigg O, Kinsman R, Treasure T. Risk-adjusted sequential probability ratio tests: applications to Bristol, Shipman and adult cardiac surgery. *Int J Qual Health Care*. 2003;15:7-13.
20. Blackstone EH. Monitoring surgical performance. *J Thorac Cardiovasc Surg*. 2004;128:807-10.
21. Rogers CA, Reeves BC, Caputo M, Ganesh JS, Bonser RS, Angelini GD. Control chart methods for monitoring cardiac surgical performance and their interpretation. *J Thorac Cardiovasc Surg*. 2004;128:811-9.
22. Treasure T, Gallivan S, Sherlaw-Johnson C. Monitoring cardiac surgical performance: a commentary. *J Thorac Cardiovasc Surg*. 2004;128:823-5.
23. Jantz MA, McGaghie WC. It's time for a STAT assessment of bronchoscopy skills: the endobronchial ultrasound bronchoscopy (EBUS)-STAT and EBUS-transbronchial needle aspiration skill evaluation. *Am J Respir Crit Care Med*. 2012;186:703-5.
24. Kemp SV, El Batrawy SH, Harrison RN, Skwarski K, Munavvar M, Rosell A, et al. Learning curves for endobronchial ultrasound using cusum analysis. *Thorax*. 2010;65:534-8.
25. Spiegelhalter DJ. Monitoring clinical performance: a commentary. *J Thorac Cardiovasc Surg*. 2004;128:820-2.
26. Detterbeck F, Puchalski J, Rubinowitz A, Cheng D. Classification of the thoroughness of mediastinal staging of lung cancer. *Chest*. 2010;137:436-42.

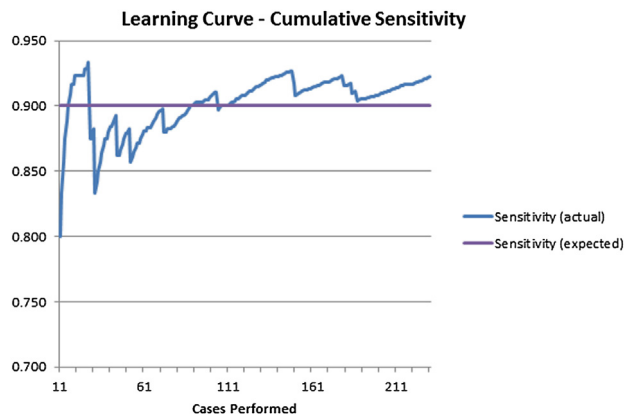


FIGURE E1. Observed sensitivity of endobronchial ultrasound-guided transbronchial needle aspiration on an institution level compared with an expected sensitivity of 90%.

GTS